Noisy channel coding via privacy amplification and information reconciliation

Joseph M. Renes and Renato Renner

Abstract—We show that optimal protocols for noisy channel coding of public or private information over either classical or quantum channels can be directly constructed from two more primitive information-theoretic tools: privacy amplification and information reconciliation, also known as data compression with side information. We do this in the one-shot scenario of structure-less resources, and formulate our results in terms of the smooth min- and max-entropy. In the context of classical information theory, this shows that essentially all two-terminal protocols can be reduced to these two primitives, which are in turn governed by the smooth min- and max-entropies, respectively. In the context of quantum information theory, the recently-established duality of these two protocols means essentially all two-terminal protocols can be constructed using just a single primitive.

Index Terms—quantum information, channel coding, privacy amplification, information reconciliation, Slepian-Wolf coding, smooth entropies

O NE of the major trends in information theory, both classical and quantum, is that a small set of proof techniques can be used to construct a wide variety of protocols. Random coding is as ubiquitous as it is useful in classical information theory, and the method of decoupling increasingly plays a similar role in quantum information theory. Instead of reusing proofs, a different approach is to reuse the protocols themselves, building up more complicated protocols by combining simpler ones. The goal is to do this in such a way that the inner workings of the parts do not have to be analyzed to ensure the correct functioning of the overall protocol. For instance, joint source-channel coding can be accomplished by simply combining a data compressor with a channel coding scheme [1]. In the quantum realm, the "mother of all" protocols, a fully-quantum version of the Slepian-Wolf task, can generate a variety of two terminal protocols involving entanglement when combined with teleportation and dense coding [2].

In this paper we construct optimal protocols for communication of classical information over noisy channels from two simpler primitives: randomness extraction and information reconciliation, also known as data compression with side information. The construction works for either classical channels or quantum channels explicitly accepting classical inputs, and by replacing randomness extraction with privacy amplification, we directly obtain a protocol for private communication.

J.M. Renes is with the Institut für Angewandte Physik, Technische Universität Darmstadt, Hochschulstr. 4a, 64289 Darmstadt, Germany. Email: joe.renes@physik.tu-darmstadt.de.

R. Renner is with the Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland. E-mail: renner@phys.ethz.ch.

We work in the one-shot scenario of structureless resources, meaning the coding scheme does not rely on repeated uses of a memoryless channel. Rather, the one-shot scenario is considerably broader in approach, encompassing not only channels in the traditional sense of communication (both with and without memory), but also channels as models for the dynamics of a physical system, for which the memoryless assumption would be out of place.

Besides adopting a new technique to construct the protocols, the resulting capacity expressions are novel as well. We find that the capacities of a channel for one-shot public [1] or private communication can be characterized in terms of smooth conditional min- and max-entropies, introduced and characterized in [3], [4], [5]. Furthermore, these expressions are shown to be essentially tight, up to small additive terms. Appealing to the asymptotic equipartition property (AEP) for smooth entropies [6] allows us to quickly recover the usual capacity expressions in the memoryless case, from Shannon's original result on the capacity of the classical channel for public communication [7] and the associated capacity for private communication [8], [9], [10], to the capacity of a quantum channel for public classical communication (known colloquially as the Holevo-Schumacher-Westmoreland, or HSW, Theorem) [11], [12] as well as for private communication [13]. Furthermore, dividing the problem of noisy channel communication into questions of coding and questions of channel properties considerably simplifies the logical arguments and should be of independent pedagogical value.

One-shot expressions for the capacity of public communication have been derived before. In [14] the one-shot capacity of a classical channel was characterized in terms of smooth min- and max-entropies, while [15] derives an expression for the capacity of quantum channels in terms of generalized (Rényi) relative entropies following a hypothesis-testing approach. These results can be seen as generalizations of earlier (asymptotic) results based on the information spectrum method [16], [17]. Very recently, [18] finds tight bounds on the capacity in terms of a smooth relative entropy quantity again from a hypothesis-testing approach. Combining the latter results with those here implies a relation between the smooth relative and conditional entropies.

Both of the primitive tasks used here are designed to manipulate "static" resources, in the sense that the goal is to transform randomness shared by distant parties into a different form (so that the joint distribution of the values held by the

---

[1]We use 'public communication' to refer to the usual task of sending classical information, in order to better distinguish it from the case of private communication.

parties is close to a given one). This task should be performed using local operations and a limited amount of communication. In particular, randomness extraction corresponds to the task of generating uniformly-distributed random variables out of non-uniform inputs, while information reconciliation uses classical communication to correlate, or reconcile, a random variable held by one party (Alice) with that held by another (Bob). One can view the classical data transmitted for this latter task as a compression of Alice's random variable, as it can be decompressed by Bob with the help of his random variable (or quantum system).

Intuitively, it is plausible that the static information reconciliation protocol could be adapted to enable reliable communication over a noisy channel, a "dynamic" resource, in the following manner, depicted schematically in Figure 1. Assuming uniform distribution of the channel inputs, consider the random variables describing the input X and output Y (where the latter may be quantum in the case of a quantum channel). As described above, information reconciliation enables Bob to reconstruct X from the compressed version of it, C, along with his information Y. Now suppose Alice and Bob agree on a particular $C = c^*$ in advance for the communication task, in that Alice restricts her channel inputs X to those with compressed output c^*. Upon receipt of Y, Bob can reconstruct X by simply reusing the decompressor of the information reconciliation protocol. In this way, each information reconciliation protocol defines a channel coding scheme: every value $C = c$ specifies a channel code consisting of all the possible inputs which compress to that value c. Since we assumed uniform distribution of the channel inputs, this coding scheme is generally not optimal. However, by running a randomness extractor backwards we can create the optimal input distribution from a uniform one, circumventing this problem. This is similar to a method used by Gallager [19] and later expanded in [20], [21].

The remainder of the paper is devoted to making this intuition rigorous. We begin in the next section by formally specifying the problem and stating the results in Theorem 1. We then move immediately to the proof of the direct part, achievability, in Section II and the converse in Section III. In Section IV we show how the usual results may be quickly recovered for the case of very many uses of a memoryless channel. We conclude in Section V by discussing applications of this result and its relation to other work.

I. DEFINITIONS AND RESULTS

We work directly with a classical-quantum channel Θ taking input classical symbols $x \in \mathcal{X}$ to output quantum states $\vartheta_x^Y \in \mathcal{S}(\mathcal{H}^Y) \equiv \mathcal{Y}$. To recover the case of a classical channel, one can simply require that the output states ϑ_x^Y be simultaneously diagonalizable.

We note that the restriction to classical channel inputs can be made without loss of generality as we are interested in transmitting classical information, either publicly or privately. In terms of a physical channel accepting quantum inputs, this just amounts to fixing the quantum states to be input for given classical value x; here this choice is effectively part of the channel (this is possible since, in the one-shot treatment adopted here, the channel is only used once). On the other hand, one may regard this choice as part of the encoder, and the only necessary modification of the expression for the capacity (see Theorem 1 below) would be to include an optimization over this choice.

An (n, ε)-coding scheme for a classical-quantum channel consists of an encoder $\mathsf{Enc} : \mathcal{M} \to \mathcal{X}$ taking classical messages $m \in \mathcal{M}$ to channel inputs and a decoder $\mathsf{Dec} : \mathcal{Y} \to \mathcal{M}$ taking channel outputs to guesses of the input messages, for which $n = \log_2 |\mathcal{M}|$ and

$$p_{\text{error}}(m) \equiv \Pr[m \neq \mathsf{Dec} \circ \Theta \circ \mathsf{Enc}(m)] \leq \varepsilon \qquad (1)$$

for all $m \in \mathcal{M}$. In addition, if Θ outputs a bipartite state ϑ^{YZ}, of which Bob receives the Y subsystem and an eavesdropper Eve the Z subsystem living in $\mathcal{S}(\mathcal{H}^Z) \equiv \mathcal{Z}$, then an (n, ε)-private coding scheme is an encoder-decoder pair as above, with the additional requirement that every message m be approximately unknown to the eavesdropper:

$$p_{\text{secret}}(m) \equiv \tfrac{1}{2} \left\| \vartheta_{\mathsf{Enc}(m)}^Z - \vartheta^Z \right\|_1 \leq \varepsilon, \qquad (2)$$

where $\vartheta^Z = \frac{1}{2^n} \sum_m \vartheta_{\mathsf{Enc}(m)}^Z$.[2]

It is useful to think of the message as being a random variable M, taking values in $\mathcal{M}$ according to the probability distribution P_M. Then the message transmission process is encapsulated by the following sequence of random variables (Markov chain). We use a prime to denote a random variable or quantum system which is meant to be nearly identical to the unprimed version; in the present context we would like the output M' to be essentially equal to the input M.

$$M \xrightarrow{\mathsf{Enc}} X \xrightarrow{\Theta} Y \xrightarrow{\mathsf{Dec}} M' \qquad (3)$$

Given a choice of ε, the capacity for public communication (usually just referred to as the classical capacity) $\mathcal{C}_{\text{pub}}^\varepsilon(\Theta)$ of the channel is simply $\log_2 n$ for the largest n in an (n, ε) coding scheme. The private capacity $\mathcal{C}_{\text{prv}}^\varepsilon(\Theta)$ is defined similarly using private coding schemes. Here we prove the following upper and lower bounds on these capacities in terms of the smooth min- and max-entropies, which are defined in the appendix.

Theorem 1 (Capacities of Classical-Quantum Channels).
For all $\varepsilon > 0$,

$$C_{\text{pub}}^\varepsilon(\Theta) \geq \max_{P_X} \Big[H_{\min}^{\varepsilon/8}(X) - H_{\max}^{\varepsilon/8}(X|Y)$$
$$- 4 \log \tfrac{1}{\varepsilon} - 16 \Big], \qquad (4)$$

$$C_{\text{pub}}^\varepsilon(\Theta) \leq \max_{P_X} \Big[H_{\min}(X) - H_{\max}^{\sqrt{2\varepsilon}}(X|Y) \Big]. \qquad (5)$$

For $T \to X \to (Y, Z)$ a Markov chain,

$$C_{\text{prv}}^{8\varepsilon}(\Theta) \geq \max_{P_T, T \to X} \Big[H_{\min}^{\varepsilon/8}(T|Z) - H_{\max}^{\varepsilon/8}(T|Y)$$
$$- 4 \log \tfrac{1}{\varepsilon} - 16 \Big], \qquad (6)$$

$$C_{\text{prv}}^\varepsilon(\Theta) \leq \max_{P_T, T \to X} \Big[H_{\min}^{\sqrt{2\varepsilon}}(T|Z) - H_{\max}^{\sqrt{2\varepsilon}}(T|Y) \Big]. \qquad (7)$$

---

[2]In fact, ϑ^Z need not be the average state, but can be arbitrary.

II. Achievability

The proof of the direct parts proceeds in three steps, successively building up to a construction of an encoder and decoder. The first step is to show that in doing this, we only need to worry about the average transmission error and secrecy of the communication scheme, assuming the inputs are uniformly distributed. Then, we show that protocols for information reconciliation can be adapted to the channel coding scenario, when the input to the channel is uniformly distributed (not just the messages themselves). Finally, we show how to mimic any particular channel input distribution from a uniform distribution by using randomness extraction, and how to mimic an input distribution so that the eavesdropper learns nothing about the message using privacy amplification.
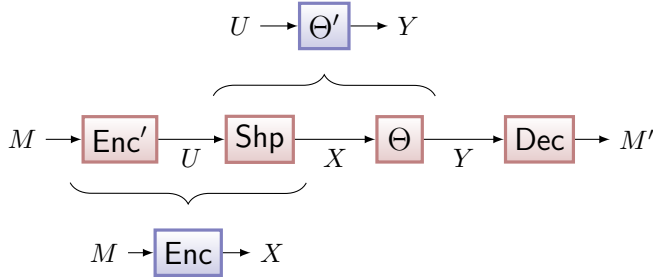

Fig. 1. Schematic of using randomness extraction and information reconciliation (data compression with side information) to perform noisy channel communication. Messages $m \in M$ are input to the encoder Enc' and subsequently to the shaper Shp, which is a randomness extractor run in reverse. Then they are then transmitted over the channel Θ to the receiver, who uses the decoder Dec to construct a guess $m' \in M'$ of the original input. Concatenating the shaper and channel gives a new effective channel Θ', for which an encoder/decoder pair $(\mathsf{Enc}', \mathsf{Dec})$ can be constructed by repurposing a compressor/decompressor pair that operates on the joint input-output UY of the channel. Ultimately, the shaper can instead be regarded as part of the encoder Enc, which is formed by concatenating Enc' and Shp.

A. Average Case Coding Implies Worst Case Coding

We start by observing that constructing an encoder/decoder pair with low *average* error probability on the receiver's end and low average trace distance of eavesdropper outputs suffices to construct an encoder/decoder pair with low error probability and secrecy parameter in the worst case.

Lemma 1 (Average Case to Worst Case Error). *Given a channel $\Theta : X \to Y$ and an encoder/decoder pair $\mathsf{Enc} : M \to X$, $\mathsf{Dec} : Y \to M'$ such that $\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} p_{\mathrm{error}}(m) \leq \frac{\varepsilon}{4}$ and $\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} p_{\mathrm{secret}}(m) \leq \frac{\varepsilon}{4}$, then there exists an encoder/decoder pair for a subset $\mathcal{M}^* \subset \mathcal{M}$ of size at least $|\mathcal{M}|/2$ such that $p_{\mathrm{error}}(m^*) \leq \varepsilon$ and $p_{\mathrm{secret}}(m^*) \leq \varepsilon$ for all $m^* \in \mathcal{M}^*$.*

Proof: By the Markov inequality, a fraction at most one-quarter of $m \in \mathcal{M}$ have $p_{\mathrm{error}}(m) \geq \varepsilon$. Similarly, for at most one-quarter does $p_{\mathrm{secret}}(m) \geq \varepsilon$. Thus, there is a subset $\mathcal{M}^*$ of half the $m \in \mathcal{M}$ for which neither statement is true. Restricting the input of Enc to $\mathcal{M}^*$ gives the new encoder. The new decoder is given by altering the old decoder so that outputs $m \notin \mathcal{M}^*$ are mapped at random to $m \in \mathcal{M}^*$. ∎

Remark 1. *If we only require that $p_{\mathrm{error}}(m) \leq \varepsilon$, then $\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} p_{\mathrm{error}}(m) \leq \frac{\varepsilon}{2}$ suffices.*

B. Channel Coding From Information Reconciliation For Uniformly-Distributed Inputs

Now we show that an information reconciliation protocol can be adapted to channel coding, at least when the input to the channel is uniformly or nearly uniformly distributed. We do this explicitly for the case of linear compression functions and subsequently remark how it can be made more general.

First we need to specify classical-quantum information reconciliation protocols more precisely. Given a classical random variable X and a quantum system Y jointly described by the classical-quantum state $\psi^{XY} = \sum_{x \in \mathcal{X}} p_x |x\rangle\langle x|^X \otimes \vartheta_x^Y$, an ε-good information reconciliation protocol consists of a compression map $\mathsf{Cmp} : \mathcal{X} \to \mathcal{C}$ taking X to another classical random variable C and a decompression map $\mathsf{Dcp} : (\mathcal{C}, \mathcal{Y}) \to \mathcal{X}$ taking C and states in the system Y to elements X' of the input alphabet $\mathcal{X}$ such that the error probability $p_{\mathrm{error}} = \sum_x p_x \mathrm{Pr}[x \neq \mathsf{Dcp}(\mathsf{Cmp}(x), \vartheta_x^Y)] \leq \varepsilon$. If the alphabet $\mathcal{X}$ forms a linear space, the compression map could be linear, and one speaks of a *linear compressor*.

Lemma 2 (Channel Coding from Information Reconciliation of Uniform Inputs). *Given a cq channel $\Theta : \mathcal{U} \to \mathcal{Y}$ from uniformly-distributed inputs U to arbitrary outputs Y, suppose the linear compressor and arbitrary decompressor pair $\mathsf{Cmp}/\mathsf{Dcp}$ form an ε-good information reconciliation protocol for the combined input and output UY. Then there exists a linear encoder $\mathsf{Enc} : \mathcal{M} \to \mathcal{U}$ and a decoder $\mathsf{Dec} : \mathcal{Y} \to \mathcal{M}$ for Θ such that the error probability of transmitting a uniformly-distributed message M of size $|\mathcal{M}| = |\mathcal{U}|/|\mathcal{C}|$ is also less than ε.*

Proof: Start by defining $p_{\mathrm{error}}(u) = \mathrm{Pr}[u \neq \mathsf{Dcp}(\mathsf{Cmp}(u), \vartheta_u)]$. Then the information reconciliation error probability can be formulated as

$$p_{\mathrm{error}} = \frac{1}{|\mathcal{U}|} \sum p_{\mathrm{error}}(u) = \frac{1}{|\mathcal{C}|} \sum_c \frac{|\mathcal{C}|}{|\mathcal{U}|} \sum_{u : \mathsf{Cmp}(u) = c} p_{\mathrm{error}}(u)$$
$$= \frac{1}{|\mathcal{C}|} \sum_c \langle p_{\mathrm{error}}(u) \rangle_{P_{U|C=c}}, \tag{8}$$

where $\langle X \rangle_{P_X}$ denotes the average of X using the distribution P_X and P_U is the uniform distribution. In other words, the average error probability is the average over outputs c of the average error probability of inputs u consistent with a given output. In this expression we have split the summation over u to first a summation over the values of c and then for each of these a summation over the u for which $\mathsf{Cmp}(u) = c$. In so doing, we have used the fact that there are $|\mathcal{U}|/|\mathcal{C}|$ preimages for each c, which follows from Lemma 3 (see appendix).

Choosing the value of $C = c^*$ with the lowest error probability $\langle p_{\mathrm{error}}(u) \rangle_{P_{U|C=c^*}}$ enables us to define an encoder and decoder from the compressor and decompressor restricted to this value. The encoder simply maps $m \in \mathcal{M}$ to those $u \in \mathcal{U}$ for which $\mathsf{Cmp}(u) = c^*$ in some fixed order, say lexicographic order. By linearity of the compressor, $|\mathcal{M}| = |\mathcal{U}|/|\mathcal{C}|$. The decoder is then defined by taking the output of the decompressor

and then applying the inverse of the encoding map to the result u' or outputting a random $m \in \mathcal{M}$ when $\mathsf{Cmp}(u') \neq c^*$.

The error probability for the encoder/decoder combination for uniformly distributed messages M is exactly the same as the error probability for the compressor/decompressor combination, which must be lower than the average by construction. ∎

Remark 2. *If the compressor/decompressor pair has error probability ε_2 when acting on a nearly uniform input U' satisfying $\frac{1}{2} \|U - U'\|_1 \leq \varepsilon_1$, then applying the corresponding encoder/decoder to a uniform input gives an error probability of at most $\varepsilon_1 + \varepsilon_2$ by the triangle inequality.*

Remark 3. *By using Lemma 4 instead of Lemma 3 (see appendix), the restriction to linear compression functions can be removed at the cost of reducing the number of messages by a factor ε and an additional failure probablity ε.*

C. Distribution Shaping

Finally, we need to remove the restriction of uniform inputs to the channel. This is done by combining the channel with a *distribution shaper*, which is a means of mapping a uniform distribution to a chosen distribution. By running the distribution shaper and then the channel, we obtain a virtual channel which acts again on a (roughly) uniformly distributed input. The distribution shaper can be constructed using a randomness extractor, as follows.

Suppose that $\mathsf{Ext} : X \to U'$ is a function which produces an ε-good approximation of a uniformly distributed random variable U from an input X distributed according to P_X, in the sense that $\frac{1}{2} \|U - U'\|_1 \leq \varepsilon$. The extractor defines a joint distribution $P_{XU'}$, and with this we can define a function $\mathsf{Shp} : (U', R) \to X$ which is in some sense the inverse of Ext. Here R is some additional randomness, and Shp is defined by using R to select an $x \in X$ from the distribution $P_{X|U'=u'}$ given the input value $U' = u'$. Thus, the output of the shaper is again X. Shapers constructed in this manner will be called ε-shapers. Moreover, if the extractor performs privacy amplification of X against some Z generated from X, then the shaper replicates X while hiding U' from the eavesdropper. This follows because the conditional states relevant to the eavesdropper are the same in both cases.

It may seem strange to additionally require a source of randomness for this purpose, and ideally we would like all the randomness needed to generate X to be contained in U'. However, the mapping that takes a general X to a nearly-uniform distribution U' may map two values x to the same u'. When that u' is input to Shp, some randomness is needed to reverse the mapping.

D. Putting it all together

Now we can combine these three pieces to establish the direct part of Theorem 1. We do this first for the private capacity and then make some modifications to obtain the lower bound on the classical capacity. The latter can be obtained as a special case of the former by assuming the channel does not

leak anything to an adversary, i.e. Z is trivial, but the additional modifications will improve the constants in the bound.

For a given channel $\Theta : X \to Y$ and input distribution P_X, we can define a new channel $\Theta' : U' \to Y$ (with output states $\vartheta_{u'}^{'Y}$) by concatenating an ε_1-shaper Shp that generates X, built from a privacy amplification extractor, with Θ and regarding R as part of the channel. Next, following Remark 2 we construct an encoder/decoder pair $\mathsf{Enc}'/\mathsf{Dec}'$ from an ε_2-good compressor/decompressor for $\psi^{U'Y} = \sum_{u' \in \mathcal{U}} p_{u'} |u'\rangle\langle u'|^{U'} \otimes \vartheta_{u'}^{'Y}$, where $p_{u'}$ is the distribution of the input U' to the shaper Shp. When input with a uniformly distributed U, the error probability averaged over codewords and choices of code is at most $\varepsilon_1 + \varepsilon_2$, while the average leakage to the eavesdropper is at most $2\varepsilon_1$. For simplicity, define $\varepsilon \equiv 4\max(\varepsilon_1, \varepsilon_2)$. By the Markov inequality, at least three-quarters of the code choices have an average codeword error rate below $4(\varepsilon_1 + \varepsilon_2) < 2\varepsilon$. By the same reasoning, at least three-quarters of the code choices have an average p_{secret} less than $8\varepsilon_1 < 2\varepsilon$. Therefore, at least half have both properties.

Regarding the shaper as part of the encoder instead of part of the channel, we can define $\mathsf{Enc} = \mathsf{Shp} \circ \mathsf{Enc}'$. Applying Lemma 1, we can then make the further adjustments to Enc and Dec to simultaneously achieve a worst-case error of 8ε and worse-case leakage 8ε.

Finally, we can count how many messages can be reliably sent using the constructed encoder and decoder. From Lemma 2, we have $n \geq \log|\mathcal{U}| - \log|\mathcal{C}| - 1$. Inserting the known results for privacy amplification and data compression, Theorems 2 and 3 in the appendix, this becomes

$$n \geq H_{\min}^{\varepsilon_{11}}(X|Z) - 2\log\frac{1}{\varepsilon_{12}} - H_{\max}^{\varepsilon_{21}}(U'|Y) - 2\log\frac{1}{\varepsilon_{22}} - 4, \tag{9}$$

where $\varepsilon_1 = \varepsilon_{11} + \varepsilon_{12}$ and $\varepsilon_2 = \varepsilon_{21} + \varepsilon_{22}$. Again for simplicity, let $\varepsilon_{jk} = \varepsilon/8$. Because U' is a function of X via the extractor, the max-entropy cannot increase when replacing X by U'. Since we are free to choose any P_X in this argument we therefore have

$$n \geq \max_{P_X} \left[H_{\min}^{\varepsilon/8}(X|Z) - H_{\max}^{\varepsilon/8}(X|Y) - 4\log\frac{1}{\varepsilon} - 16 \right]. \tag{10}$$

To complete the argument for the private capacity, note that Alice could precede the channel with another mapping from T to X, which she is free to optimize. Regarding this as part of the original channel in the above argument then leads to the desired result.

The direct part for the channel capacity follows by making a few small modifications. First, the Markov inequality is no longer needed to ensure the two conditions of private communication are satisfied. Here there is only one, and certainly there exists an encoding with average codeword error probability less than the average over codes and codewords, $\varepsilon_1 + \varepsilon_2$. We then only require Remark 1 rather than Lemma 1 to move to the worst-case error $2(\varepsilon_1 + \varepsilon_2)$ over codewords. Finally, though in principle it is also possible for Alice to precede the channel with a $T \to X$ mapping, we shall see in the converse that this is not necessary. Note also that in this context the encoder can dispense with the randomness needed

to properly simulate X and just fix a particular value of the output, for instance the x with the largest $P_{X=x|C=c^*}$.

III. CONVERSE

We first prove the converse for the private capacity and then modify the argument to establish the converse for the classical capacity. Given an (n, ε)-private coding scheme, the two requirements of the output made by the definition imply that $H_{\max}^{\sqrt{2\varepsilon}}(M|M') \leq 0$ and $H_{\min}^{\sqrt{2\varepsilon}}(M|Z) \geq n$, the former for any distribution of messages and the latter for the uniform distribution. The former follows because the trace distance of the pair (M, M') to (M, M) is less than ε and therefore (M, M) is in the $\sqrt{2\varepsilon}$-neighborhood of (M, M') (see the appendix; the square root is a consequence of the conversion of the trace to purification distance). But $H_{\max}(M|M) = 0$ and thus $H_{\max}^{\sqrt{2\varepsilon}}(M|M') \leq 0$. The latter follows because again the ideal output, in which Z is independent of the uniformly-distributed M and therefore satisfies $H_{\min}(M|Z) \geq n$, is in the $\sqrt{2\varepsilon}$-neighborhood of the actual pair (M, Z). Additionally, by the data processing inequality [5], $H_{\max}^{\sqrt{2\varepsilon}}(M|M') \geq H_{\max}^{\sqrt{2\varepsilon}}(M|Y)$ since the decoder generates the guess M' from Y.

Defining $\bar{P}_M$ to be the uniform distribution, we have

$$
\begin{aligned}
\max_{P_M, M \to X} & \left[H_{\min}^{\sqrt{2\varepsilon}}(M|Z) - H_{\max}^{\sqrt{2\varepsilon}}(M|Y) \right] \\
&\geq \max_{M \to X} \left[H_{\min}^{\sqrt{2\varepsilon}}(M|Z)_{\bar{P}_M} - H_{\max}^{\sqrt{2\varepsilon}}(M|Y)_{\bar{P}_M} \right] \\
&\geq \max_{M \to X} \left[H_{\min}^{\sqrt{2\varepsilon}}(M|Z)_{\bar{P}_M} - H_{\max}^{\sqrt{2\varepsilon}}(M|M')_{\bar{P}_M} \right] \\
&\geq n,
\end{aligned}
$$

which is the form we set out to prove.

For the converse of the classical capacity, observe that the encoding function Enc is without loss of generality deterministic and injective. It might as well be deterministic, since if it used randomness, we could make it deterministic by fixing the randomness to that value with the least probability of error, which cannot be worse than the average case. Moreover, for this deterministic choice, Enc must be injective, since a collision of two inputs having the same codeword necessarily implies an error. Now, using the injectivity of Enc we can define a distribution $\bar{P}_X$ given a distribution over M by simply taking the distribution of M on its image in X and zero otherwise. Choosing the uniform distribution over M and observing that $H_{\min}(M) = n$ when M is uniformly distributed, we obtain

$$
\max_{P_X} \left[H_{\min}(X) - H_{\max}^{\sqrt{2\varepsilon}}(X|Y) \right] \tag{11}
$$

$$
\geq \left[H_{\min}(X)_{\bar{P}_X} - H_{\max}^{\sqrt{2\varepsilon}}(X|M')_{\bar{P}_X} \right] \tag{12}
$$

$$
= \left[H_{\min}(M) - H_{\max}^{\sqrt{2\varepsilon}}(M|M') \right] \tag{13}
$$

$$
\geq n. \tag{14}
$$

IV. ASYMPTOTIC ANALYSIS

In the asymptotic limit of $n \to \infty$ uses of a memoryless channel we recover the known results on the *rate* of public or private communication, where the rate of private communication is defined by

$$
R_{\mathrm{prv}}(\Theta) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\mathcal{C}_{\mathrm{prv}}^{\varepsilon}(\Theta^{\otimes n})}{n}, \tag{15}
$$

and the rate of public communication is defined similarly. In general, the rates take the rather ungainly form

$$
R_{\mathrm{pub}}(\Theta) = \lim_{\ell \to \infty} \frac{1}{\ell} \max_{P_{X^\ell}} \left[H(X^\ell) - H(X^\ell|Y^{\otimes \ell}) \right], \tag{16}
$$

$$
R_{\mathrm{prv}}(\Theta) = \lim_{\ell \to \infty} \frac{1}{\ell} \max_{P_T, T \to X^\ell} \left[H(T|Z^{\otimes \ell}) - H(T|Y^{\otimes \ell}) \right]. \tag{17}
$$

Here X^n refers to a classical random variable on $\mathcal{X}^n$, while $Y^{\otimes n}$ refers to the n-fold tensor product of the Hilbert space $\mathcal{H}^Y$ and similarly for Z. The rate for public communication over quantum channels is known as the HSW theorem, after Holevo [11] and Schumacher and Westmoreland [12]. The private rate was proven by Devetak [13].

For the special case of classical channel outputs, i.e. ϑ_x^Y (and separately $\widetilde{\vartheta}_x^Z$) are all simultaneously diagonalizeable, these reduce to the familiar and simpler form

$$
R_{\mathrm{pub}}(\Theta_{\mathrm{cl}}) = \max_{P_X} \left[H(X) - H(X|Y) \right], \tag{18}
$$

$$
R_{\mathrm{prv}}(\Theta_{\mathrm{cl}}) = \max_{P_T, T \to X} \left[H(T|Z) - H(T|Y) \right]. \tag{19}
$$

The classical rate is Shannon's original noisy channel coding theorem [7]. The private rate was first established by Wyner in the specific setting of the wire-tap channel [8], later expanded to arbitrary channels by Ahlswede and Csiszar [9], and strengthened to the stronger form of security used here (cf. Eq. 2) by Maurer and Wolf [10].

The proof proceeds in two steps. First we show that such rates are possible using the lower bound on the capacity and applying the asymptotic equipartition property (AEP) of the conditional min- and max-entropies. Then we show that the rates cannot be exceeded by making use of the upper bound on the capacity and bounds on the conditional min- and max-entropy in terms of the conditional von Neumann entropy. Since the case of public communication follows from that of private communication, we only give the argument for the latter.

For the direct part, we begin with the lower bound on the capacity from Theorem 1. For m uses of the channel Θ, this becomes

$$
\mathcal{C}_{\mathrm{prv}}^{8\varepsilon}(\Theta^{\otimes m}) \geq \max_{P_T, T \to X^m} \left[H_{\min}^{\varepsilon/8}(T|Z^{\otimes m}) - H_{\max}^{\varepsilon/8}(T|Y^{\otimes m}) \right.
$$
$$
\left. - 4\log\frac{1}{\varepsilon} - 16 \right]. \tag{20}
$$

Since this is a lower bound, we're free to choose T as we like. We choose $T = T^m$ to be i.i.d., each instance T_i separately generating the likewise i.i.d. X_i via some fixed map $T \to X$. Now we make use of the AEP, which states [6]

$$
\lim_{\varepsilon \to 0} \lim_{m \to \infty} H_{\min}^{\varepsilon}(X^m|Z^{\otimes m}) = mH(X|Z), \tag{21}
$$

and similarly for the conditional max-entropy. We then obtain

$$\lim_{\varepsilon\to 0}\lim_{m\to\infty}\frac{1}{m}\mathcal{C}_{\mathrm{prv}}^{8\varepsilon}(\Theta^{\otimes m}) \geq \max_{P_T, T\to X}\left[H(T|Z)-H(T|Y)\right].\tag{22}$$

Finally, we let $n=\ell m$, and replay the above argument using the superchannel $\Theta^{\otimes\ell}$ (ℓ independent uses of the channel) to obtain the desired result.

Note that, in contrast to the standard proof technique, here we only need to make a statement about the entropies in the capacity formula, a statement provided by the AEP. Importantly, typical sequences, type classes, or the like play no role in the protocol itself, but could be used to establish the AEP. Such methods are not necessary; indeed, the approach of [6] is based on properties of Rényi entropies.

To complete the argument, we consider the upper bound on the capacity, and use the bounds on the conditional min- and max-entropies from Lemma 5, replacing $\dim(A)$ with $|\mathcal{X}|$. Now the upper bound on the private capacity becomes

$$\mathcal{C}_{\mathrm{prv}}^{\varepsilon}(\Theta^{\otimes n}) \leq \max_{P_T, T\to X^n}\Big[H(T|Z^{\otimes n})-H(T|Y^{\otimes n})$$
$$+16n\sqrt{2\varepsilon}\log|\mathcal{X}|+4h_2(2\sqrt{2\varepsilon})\Big].\tag{23}$$

Dividing through by n and taking the limits $n\to\infty$ and $\varepsilon\to 0$ (whose order is now irrelevant) yields the desired result (replacing n by ℓ).

When the channel has purely classical outputs, the limit involving $\ell\to\infty$ (called regularization) can be removed. For private communication we show this explicitly in Lemma 6, which recovers Eq. 18; for public communication (Eq. 19) see, e.g. Theorem 4.2.1 in [19]. Should the channel produce quantum outputs, regularization is known to be necessary in both cases, private [22] and public [23].

Finally, we note that the optimization over maps $T\to X$ is generally necessary to achieve the optimal rate of private communication, by means of the following example. Suppose Θ is a purely classical channel defined in Fig. 2. To send private messages to Y, clearly one can encode 0 as $X=0$ or $X=1$ randomly and 1 as $X=2$. The message can be unambiguously determined from Y no matter the encoding, but Z will be completely random for either input, and so this encoding scheme achieves a rate of 1 bit. Eschewing random encoding, the maximum rate of private communication is $\max_{P_X}\left[H(X|Z)-H(X|Y)\right]$. Due to the structure of the $X\to Z$ map, the maximum of the first term is one, and this can only occur when 0 and 1 occur with equal probability. But this implies the second term is nonzero, meaning the overall rate is less than one.

V. Conclusions

Rather than reusing the proof techniques involved in understanding more basic information processing protocols such as privacy amplification and information reconciliation, here we have shown how to construct channel coding protocols from these protocols themselves. Moreover, if the underlying protocols are optimal, then so are the channel coding protocols. This provides an appealing conceptual framework for
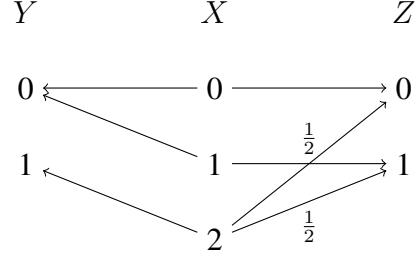

Fig. 2. Channel demonstrating the need for randomness by the encoder to achieve the private communication capacity. Unmarked arrows denote deterministic maps; otherwise the probability of a transition is marked.

two-terminal problems in information theory in which one successively builds up to more complicated protocols using simpler elements whose internal workings are not relevant for the present task. Moreover, it is also appealing to see that the two basic primitives are characterized in terms of the two basic entropic quantities, smooth min- and max-entropy, and that these quantities enter the capacity expressions in a way which reflects the protocol construction.

In the setting of quantum information theory these entropies are dual [5], as are the two primitives [24], meaning only one primitive is needed to construct more and more complicated protocols. As an example, instead of appealing to Theorem 3 for the compressor/decompressor pair needed to establish Eq. 9 in Theorem 1, we may rely on

1) Theorem 2,
2) the duality of privacy amplification and information reconciliation as shown in [24], and
3) a new form of the uncertainty principle derived in [25].

Specifically, in the proof of Theorem 1 we require a linear compressor/decompressor pair operating on the classical-quantum state of $U'Y$, classical in say the U' basis (in an abuse of notation). By the duality in [24], such a pair with error probabilty $\sqrt{2\varepsilon}$ can be constructed from ε-good linear privacy amplification of the conjugate basis $\widetilde{U}'$[3], and the size of the compressed output $\mathcal{C}$ in the optimal case is given by $\log_2|\mathcal{C}|=\log_2|\mathcal{U}|-H_{\min}^{\varepsilon_1}(\widetilde{U}'|R)+2\log\frac{1}{\varepsilon_2}-1$, where R is the purification of the original system XY and $\varepsilon=\varepsilon_1+\varepsilon_2$. But from the uncertainty principle of [25] we have

$$H_{\min}^{\varepsilon}(\widetilde{U}'|R)+H_{\max}^{\varepsilon}(U'|Y)\geq\log|\mathcal{U}|.\tag{24}$$

Therefore $\log_2|\mathcal{C}|\leq H_{\max}^{\varepsilon}(U'|Y)+O(\log\frac{1}{\varepsilon})$, and we recover Eq. 9 up to $O(\log\frac{1}{\varepsilon})$ terms.[4] Thus we have constructed all-decoupling proofs of the public and private capacities of a quantum channel, in the sense that establishling the capacity now does not rely on directly constructing a decoder for the receiver, as in the proof of Theorem 3, but rather on decoupling the purifying system, as in the proof of Theorem 2.[5] In the

[3] The cq nature of the state ensures that we satisfy condition (b) of Theorem 4 in [24].

[4] $\log_2|\mathcal{C}|$ also cannot be substantially smaller, by the lower bound of Theorem 3), and the capacity as calculated here is certainly no smaller than that of Eq. 9 (up to $O(\log\frac{1}{\varepsilon})$ terms).

[5] Observe that we could not have used a similar form of the uncertainty principle derived in [24] for the present purpose, as it subtly relies on the decoder construction we are trying to avoid.

asymptotic limit of many independent uses of the channel, we then recover the familiar HSW [11], [12] and private capacity [13] results. This derivation of the classical capacity of a quantum channel can thus be seen as the classical-quantum analogue of [26], where the quantum capacity of a quantum channel is derived using a decoupling approach. As with the main proof of Theorem 1, the decoupling procedure outlined above also suffices to derive Shannon's result on the public capacity of classical channels [7], as well as the associated privacy capacity results [8], [9], [10], simply by treating the classical channel in the formalism of quantum information theory.

APPENDIX

The conditional max-entropy for a state ρ^{AB} is defined by

$$H_{\max}(A|B)_\rho \equiv \max_{\sigma^B} \ 2 \log F(\rho^{AB}, \mathbb{1}^A \otimes \sigma^B), \qquad (25)$$

where the maximization is over positive, normalized states σ and $F(\rho,\sigma) \equiv \|\sqrt{\rho}\sqrt{\sigma}\|_1$ is the fidelity of ρ and σ. Dual to the conditional max-entropy is the conditional min-entropy,

$$H_{\min}(A|B)_\rho \equiv \max_{\sigma^B} \left(-\log \lambda_{\min}(\rho^{AB}, \sigma^B) \right), \qquad (26)$$

with $\lambda_{\min}(\rho^{AB}, \sigma^B) \equiv \min\left\{ \lambda : \rho^{AB} \le \lambda \mathbb{1}^A \otimes \sigma^B \right\}$. The two are dual in the sense that

$$H_{\max}(A|B)_\rho = -H_{\min}(A|C)_\rho \qquad (27)$$

for ρ^{ABC} a pure state [27]. The min- and max-entropies derive their names in part from the following relation (Lemma 2, [6]),

$$H_{\min}(A|B)_\rho \le H(A|B)_\rho \le H_{\max}(A|B)_\rho, \qquad (28)$$

where $H(A|B)_\rho \equiv H(AB)_\rho - H(B)_\rho$ for $H(B)_\rho = -\mathrm{Tr}[\rho \log \rho]$ is the usual conditional von Neumann entropy.

The min- and max-entropies can be *smoothed* by considering possibly subnormalized states $\bar{\rho}^{AB}$ in the ε-neighborhood of ρ^{AB}, defined using the purification distance $P(\rho,\sigma) \equiv \sqrt{1 - F(\rho,\sigma)^2}$,

$$B_\varepsilon(\rho) \equiv \{\bar{\rho} : P(\rho,\bar{\rho}) \le \varepsilon\}. \qquad (29)$$

Note that the purification distance is essentially equivalent to the trace distance, due to the bounds $D(\rho,\sigma) \le P(\rho,\sigma) \le \sqrt{2D(\rho,\sigma)}$ [5]. The smoothed entropies are then given by

$$H_{\min}^\varepsilon(A|B)_\rho \equiv \max_{\bar{\rho} \in B_\varepsilon(\rho^{AB})} H_{\min}(A|B)_{\bar{\rho}}, \qquad (30)$$

$$H_{\max}^\varepsilon(A|B)_\rho \equiv \min_{\bar{\rho} \in B_\varepsilon(\rho^{AB})} H_{\max}(A|B)_{\bar{\rho}}. \qquad (31)$$

Furthermore, the dual of $H_{\max}^\varepsilon(A|B)_\rho$ is $H_{\min}^\varepsilon(A|C)_\rho$, so that taking the dual and smoothing can be performed in either order [5].

Optimal one-shot privacy amplification results are established in [4], [28], [29], [30]. Using the entropy definitions above, the number of ε-good random bits $\ell_{\mathrm{ext}}^\varepsilon(X|E)$ which can be extracted from the classical random variable X against a possibly quantum eavesdropper is bounded by

Theorem 2 (Privacy Amplification). *Given a state* $\psi^{XE} = \sum_x p_x |x\rangle\langle x|^X \otimes \varphi_x^E$ *and* $\varepsilon_1, \varepsilon_2 \ge 0$ *such that* $\varepsilon = \varepsilon_1 + \varepsilon_2$,

$$H_{\min}^{\varepsilon_1}(X|E)_\psi - 2\log\frac{1}{\varepsilon_2} + 1 \le \ell_{\mathrm{ext}}^\varepsilon(X|E)_\psi \le H_{\min}^{\sqrt{2\varepsilon}}(X|E)_\psi.$$

Meanwhile, optimal one-shot information reconciliation results are given in [31]. The minimum number of bits $\ell_{\mathrm{cmp}}^\varepsilon(X|B)$ to which the classical random variable X can be compressed and still be recovered using side information B at the decoder with error probability less than ε is bounded by

Theorem 3 (Classical-Quantum Information Reconciliation). *Given a state* $\psi^{XB} = \sum_x p_x |x\rangle\langle x|^X \otimes \varphi_x^B$ *and* $\varepsilon_1, \varepsilon_2 \ge 0$ *such that* $\varepsilon = \varepsilon_1 + \varepsilon_2$

$$H_{\max}^{\sqrt{2\varepsilon}}(X|B)_\psi \le \ell_{\mathrm{cmp}}^\varepsilon(X|B)_\psi \le H_{\max}^{\varepsilon_1}(X|B)_\psi + 2\log\frac{1}{\varepsilon_2} + 4.$$

Lemma 3 (Preimage Sizes of Linear Functions). *Let* $f : \mathcal{X} \to \mathcal{Y}$ *be a linear function. Then* $|f^{-1}(y)| = |\mathcal{X}|/|\mathcal{Y}|$ *for all* $y \in \mathcal{Y}$.

Proof: Pick an x^* and consider all the $x_j \in \mathcal{X}$ such that $f(x_j) = f(x^*)$. Forming the differences $w_j = x_j - x^*$, it follows from linearity that $f(w_j) = 0$ for all j. Now consider an arbitrary $x' \in \mathcal{X}$. Clearly $f(x' + w_j) = f(x')$ for all j, so each output value has the same number of preimages. ∎

Lemma 4 (Preimage Sizes of Arbitrary Functions). *Let* $f : \mathcal{X} \to \mathcal{Y}$ *be an arbitrary function and denote by* $\mathcal{X}_y$ *the preimage of an output* y. *For a randomly-chosen output value* y, $|\mathcal{X}_y| \ge \varepsilon|\mathcal{X}|/|\mathcal{Y}|$ *with probability at least* $1 - \varepsilon$.

Proof: Let X be a uniform random variable over $\mathcal{X}$ and $Y = f(X)$. By the min-entropy chain rule (Lemma 3.1.10 in [4]) we have

$$H_{\min}(X|Y) \ge H_{\min}(XY) - \log|\mathrm{supp}(P_Y)|$$
$$= H_{\min}(X) - \log|\mathrm{supp}(P_Y)| \ge \log|\mathcal{X}|/|\mathcal{Y}|. \qquad (32)$$

Here $|\mathrm{supp}(P_Y)|$ is the size of the support of the distribution P_Y, the number of values taking nonzero probability. By the normalization condition for $P_{X|Y=y}$ it follows that $1/|\mathcal{X}_y| \le \max_x P_{X|Y=y}(x)$. And by the definition of $H_{\min}(X|Y)$,

$$\sum_{y \in \mathcal{Y}} P_{Y=y} \frac{1}{|\mathcal{X}_y|} \le 2^{-H_{\min}(X|Y)} \le \frac{|\mathcal{Y}|}{|\mathcal{X}|}. \qquad (33)$$

Finally, applying the Markov inequality to the random variable $|\mathcal{X}_Y|$ yields

$$\Pr\left[\frac{1}{|\mathcal{X}_y|} \ge \frac{|\mathcal{Y}|}{\varepsilon|\mathcal{X}|}\right] \le \frac{\varepsilon|\mathcal{X}|}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} P_{Y=y} \frac{1}{|\mathcal{X}_y|} \le \varepsilon, \qquad (34)$$

which concludes the proof. ∎

Lemma 5 (Smooth Entropy Bounds).

$$H_{\min}^\varepsilon(A|B)_\rho \le H(A|B)_\rho + 8\varepsilon \log \dim(A) + 2h_2(2\varepsilon) \quad (35)$$

$$H_{\max}^\varepsilon(A|B)_\rho \ge H(A|B)_\rho - 8\varepsilon \log \dim(A) - 2h_2(2\varepsilon), \qquad (36)$$

where $h_2(x) = -x \log_2 x - (1-x) \log_2 (1-x)$ is the binary entropy function.

Proof: Let $\bar{\rho}$ be a state in $B_\varepsilon(\rho)$ such that $H_{\min}(A|B)_{\bar{\rho}} = H_{\min}^\varepsilon(A|B)_\rho$. Then from Eq. 28 we have $H_{\min}^\varepsilon(A|B)_\rho \leq H(A|B)_{\bar{\rho}}$. Since the purification distance bounds the trace distance, $D(\rho, \bar{\rho}) \leq \varepsilon$ and we can use the continuity of the conditional von Neumann entropy [32] to establish that $H(A|B)_{\bar{\rho}} \leq H(A|B)_\rho + 8\varepsilon \log \dim(A) + 2h_2(2\varepsilon)$, completing the proof for the min-entropy. An entirely similar argument holds for the max-entropy. ∎

Lemma 6 (Single-Letter Formula for the Private Capacity)**.** *For a channel Θ_{cl} with purely classical outputs Y and Z,*

$$R_{\mathrm{prv}}(\Theta_{\mathrm{cl}}) \leq \max_{P_T, T \to X} \left[H(T|Z) - H(T|Y) \right]. \qquad (37)$$

Proof: Start with the expression $b = H(T|Z^{\otimes \ell}) - H(T|Y^{\otimes \ell})$ from Eq. 17. By Lemma 4.1 of [9], or direct calculation, we can rewrite this as a sum

$$b = \sum_{i=1}^{\ell} b_i \equiv \sum_{i=1}^{\ell} H(T|V_i Z_i) - H(T|V_i Y_i) \qquad (38)$$

for $V_i = Z_1 \ldots Z_{i-1} Y_{i+1} \ldots Y_\ell$. Observe that the random variables involved in b_i form the Markov chain $(T, V_i) \leftrightarrow X_i \leftrightarrow (Y_i, Z_i)$. This follows because $(X_i, X', Y', Z') \leftrightarrow X_i \leftrightarrow (Y_i, Z_i)$ is a Markov chain, where X' denotes the tuple of X_j random variables omitting X_i, and (T, V_i) can be computed from (X_i, X', Y', Z').

But now we can maximize each term over (T, V_i) subject to the Markov chain condition and obtain the single-letter formula

$$b \leq \sum_{i=1}^{\ell} \max_{(T,V_i)} \left[H(T|V_i Z_i) - H(T|V_i Y_i) \right]$$
$$\leq \max_{(T,V)} \ell \left[H(T|VZ) - H(T|VY) \right],$$

for the Markov chain $(T, V) \leftrightarrow X \leftrightarrow (Y, Z)$. Since Shannon entropies conditioned on V are averages of entropies conditioned on specific values $V = v$, this implies

$$b \leq \max_{(T,V,v)} \ell \left[H(T|Z, V=v) - H(T|Y, V=v) \right].$$

Finally, conditioning on $V = v$ preserves the Markov chain $T \leftrightarrow X \leftrightarrow (Y, Z)$, since $P_{YZ|X} = P_{YZ|XTV}$ implies $P_{YZ|X,V=v} = P_{YZ|XT,V=v}$. And because each choice of V and v induces a conditional distribution $P_{T|V=v}$, the maximization need only be taken over T. Using this in Eq. 17 completes the proof. ∎

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*, 2nd ed. Wiley-Interscience, Jul. 2006.

[2] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, "The mother of all protocols: restructuring quantum informations family tree," *Proc. R. Soc. A*, vol. 465, no. 2108, pp. 2537–2563, Aug. 2009.

[3] R. Renner and S. Wolf, "Smooth renyi entropy and applications," in *Proceedings of the 2004 International Symposium on Information Theory*, 2004, p. 233.

[4] R. Renner, "Security of quantum key distribution," Ph.D. dissertation, ETH Zürich, Sep. 2005.

[5] M. Tomamichel, R. Colbeck, and R. Renner, "Duality between smooth Min- and Max-Entropies," *IEEE Trans. on Inf. Theory*, vol. 56, pp. 4674-4681, 2010.

[6] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5840–5847, Nov. 2009.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[8] A. D. Wyner, "The Wire-Tap channel," *Bell Sys. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[9] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. i. secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[10] U. Maurer and S. Wolf, "Information-Theoretic key agreement: From weak to strong secrecy for free," ser. Lecture Notes on Computer Science, B. Preneel, Ed., vol. 1807. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

[11] A. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 269–273, 1998.

[12] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, p. 131, Jul. 1997.

[13] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 44– 55, 2005.

[14] R. Renner, S. Wolf, and J. Wullschleger, "The Single-Serving channel capacity," in *Information Theory, 2006 IEEE International Symposium on*, 2006, pp. 1424–1427.

[15] M. Mosonyi and N. Datta, "Generalized relative entropies and the capacity of classical-quantum channels," *Journal of Mathematical Physics*, vol. 50, no. 7, pp. 072104–14, Jul. 2009.

[16] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 7, pp. 1753–1768, 2003.

[17] G. Bowen and N. Datta, "Quantum coding theorems for arbitrary sources, channels and entanglement resources," *arXiv:quant-ph/0610003*, Sep. 2006.

[18] L. Wang and R. Renner, "One-Shot Classical-Quantum capacity and hypothesis testing," *arXiv: 1007.5456*, Jul. 2010.

[19] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, Jan. 1968.

[20] X. Ma, N. Varnica, and A. Kavčić, "Matched information rate codes for binary ISI channels," in *Proceedings of the 2002 IEEE International Symposium on Information Theory*, 2002, p. 269.

[21] J. Soriaga and P. Siegel, "On distribution shaping codes for partial-response channels," in *Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, Oct. 2003, pp. 468–477.

[22] G. Smith, J. M. Renes, and J. A. Smolin, "Structured codes improve the Bennett-Brassard-84 quantum key rate," *Phys. Rev. Lett.*, vol. 100, no. 17, pp. 170502–4, May 2008.

[23] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nat Phys*, vol. 5, no. 4, pp. 255–257, Apr. 2009.

[24] J. M. Renes, "Duality of privacy amplification against quantum adversaries and data compression with quantum side information," *Proc. Roy. Soc. A*, published online before print Dec. 2010; doi:10.1098/rspa.2010.0445

[25] M. Tomamichel and R. Renner, "The uncertainty relation for smooth entropies," *arXiv: 1009.2015*, Sep. 2010.

[26] P. Hayden, M. Horodecki, A. Winter, and J. Yard, "A decoupling approach to the quantum capacity," *Open Systems & Information Dynamics*, vol. 15, no. 1, pp. 7–19, Mar. 2008.

[27] R. König, R. Renner, and C. Schaffner, "The operational meaning of min- and Max-Entropy," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4337–4347, 2009.

[28] R. Renner and R. König, "Universally composable privacy amplification against quantum adversaries," in *Second Theory of Cryptography Conference*, ser. Lecture Notes in Computer Science, vol. 3378/2005. Cambridge, MA: Springer, Feb. 2005, pp. 407–425.

[29] R. König and R. Renner, "Sampling of min-entropy relative to quantum knowledge," *arXiv: 0712.4291*, Dec. 2007.

[30] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, "Leftover hashing against quantum side information," *arXiv: 1002.2436*, Feb. 2010.

[31] J. M. Renes and R. Renner, "One-Shot classical data compression with quantum side information and the distillation of common randomness or secret keys," *arXiv: 1008.0452*, Aug. 2010.

[32] R. Alicki and M. Fannes, "Continuity of quantum conditional informa-
tion," *Journal of Physics A: Mathematical and General*, vol. 37, no. 5,
pp. L55–L57, 2004.